Data smoothing in Fuzzy Entropy-based Battery State of Health Estimation

Sui, Xin; He, Shan; Huang, Xinrong; Teodorescu, Remus; Stroe, Daniel-Ioan

Published in:
IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society

DOI (link to publication from Publisher):
10.1109/IECON43393.2020.9255281

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Data smoothing in Fuzzy Entropy-based Battery State of Health Estimation

Xin Sui, Shan He, Xinrong Huang, Remus Teodorescu, Daniel-Ioan Stroe
Department of Energy Technology
Aalborg University
Aalborg, Denmark
{xin, she, hxi, ret, dis}@et.aau.dk

***Abstract*—To ensure the reliable operation of the batteries and maximize their service lifetime, it is important to have accurate knowledge of their state of health (SOH). Using data-driven methods to estimate the SOH is extensively studied and the feature data plays an important role in such methods. As fuzzy entropy (FE) can capture the variation of the voltage during the battery aging process, it can be used as a feature. In this paper, in order to reduce the noise from raw feature data, six smoothing methods are introduced to pre-process the FE. Furthermore, the relationship between the smoothed feature and SOH is established by support vector machine and Gaussian process regression. The comparison results show that adding a simply feature smoothing step before the model training can improve the SOH estimation performance. Finally, the effectiveness of the proposed method is verified by experimental results.**

Keywords—Lithium-ion battery, state of health estimation, data smoothing, fuzzy entropy, machine learning.

I. INTRODUCTION

Lithium-ion batteries have been widely applied in grid connected energy storage systems [1]. However, during long-term operation, the performance of Lithium-ion batteries is subject to degradation such as capacity fade and power decrease resulting in a limited service lifetime, which becomes an important concern for the user. . In order to guarantee the safe operation and maximize the lifetime of the battery, accurate estimation of its state of health (SOH) becomes essential [2].

Various battery SOH estimation methods have been proposed and they can be divided into three categories: the straightforward approach which is based on the charge transfer through the battery during charging or discharging, the model-based methods which combine the equivalent circuit model with the observer to realize the state estimation, and the data-driven methods which extract the degradation information contained in the measured data, and establish the potential mapping between the feature and the SOH [3]. The data-driven methods gain increasing interest because they are flexible and do not require a battery model. These methods include amongst others, support vector machine (SVM), relevance vector machine, neural networks, Gaussian process regression (GPR), etc.

As for the data-driven method, the quality of the input data is important for an accurate SOH estimation. However, the collected battery data used for SOH estimation are often subject to the different levels of noise pollution due to the inner disturbances (i.e., unknown electrochemical behavior in batteries) and the impact of environmental conditions (i.e., measurement error and stochastic load) [4]. In order to further improve the estimation accuracy, data pre-processing technique is required to remove the noise from the data [5]. In [4], a wavelet decomposition approach with different thresholds was introduced into the relevance vector machine model to reduce the uncertainty and improve the SOH estimation accuracy. In order to improve the ratio of signal to noise, Richardson et al. [6] applied the pre-smoothing step for the input data using a Savitzky-Golay filtering. As IC curves are sensitive to the noise in the V-Q curves, Li et [7] proposed a simple robust smoothing method based on Gaussian function to reduce the noise, therefore to preserve the important features related to battery aging on the curves. Li et al. [8] compared Gaussian filter and Savitzky-Golay filter, and found that Savitzky-Golay has excellent performance in capturing the peak points from the IC curve. Liang et al. [9] proposed a data smoothing method which is based on the moving average and empirical mode decomposition. Then a smoother and clearer IC curve was obtained and the features were well preserved.

In this paper, the effect of data smoothing on the performance of fuzzy entropy-based (FE-based) SOH estimation is studied, and six smoothing methods are used to pre-process the FE feature. As for comparison, two machine learning methods, i.e., SVM and GPR are adopted to establish the relationship between the aging feature (i.e., FE) of the battery and the SOH. The overview of the proposed analyzing method is illustrated in Fig. 1.

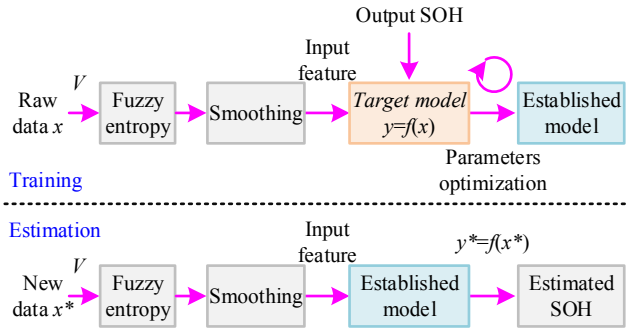

Fig. 1. Diagram of the proposed algorithm.

II. METHODOLOGY

In this section, the methods used in each link during the whole SOH estimation process are introduced, which are the FE feature extraction algorithm, six data smoothing methods, and two commonly used machine learning methods (i.e., SVM and GPR).

A. Fuzzy entropy (FE) based feature extraction method

FE is the negative natural logarithm of the conditional probability that a dataset of length N, having repeated itself for m points within a boundary, will also repeat itself for m+1 points [10]. The similarity degree is computed by the

exponential function which make FE an accurate statistic in quantifying the regularity of a data sequence. Therefore, the FE values of the voltage measured in the pulse test is used as the feature of $LiFePO_4$/C battery SOH estimation. The detail steps of the SE algorithm are shown as follows:

Step 1: Generating the $V_m(i)$ from a given feature series $\{v(1), v(2),\ldots, v(N)\}$

$$V_m(i)_{i=1\ \text{to}\ N-m+1} = [v(i), v(i+1), \ldots, v(i+m-1)] \tag{1}$$

Step 2: Removing the baseline from the generated vector $V_m(i)$, and the baseline is

$$v_0(i) = \frac{1}{m}\sum_{k=0}^{m-1} v(i+k) \tag{2}$$

Step 3: Computing the maximum distance between vectors $V_m(i)$ and $V_m(j)$, which is

$$d_{ij}^m{}_{\substack{i,j=1\ to\ N-m+1\\ k=0\ to\ m-1}} = \max\left\{\left|v(i+k)-v(j+k)\right|\right\} \tag{3}$$

Step 4: A tolerance value r is define and the similarity degree is computed using the fuzzy function

$$D_{ij}^m(r) = \exp(-\log(2)\times (d_{ij}^m / r)^2 \tag{4}$$

Step 5: Computing the conditional probability

$$B_i^m(r) = \frac{1}{N-m-1}\sum_{i,j=1}^{N-m+1} d_{ij}^m \tag{5}$$

$$A_i^m(r) = \frac{1}{N-m-1}\sum_{i,j=1}^{N-m+1} d_{ij}^{m+1} \tag{6}$$

Step 6: Defining the probability of matching points, which can be expressed as:

$$B^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} B_i^m(r) \tag{7}$$

$$A^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} A_i^m(r) \tag{8}$$

Step 7: By fixing m and r, FE can be estimated by the statistic

$$\mathrm{FE}(m,r,N) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{9}$$

Typically, the parameter m is suggested to be set at 2 or 3, and r is to be set between 0.1 and 0.25 times the standard deviation of the data [11]. In this paper, m and r is selected as 3 and 0.05, respectively.

B. Data smoothing methods

1) Moving average method (mov-average)

As shown in Fig, 2, the data are smoothed by simple averaging them in the mov-average method. The mean value is normally taken from an equal number of data on either side of a central value. For the central point x_i associated with the response value to be smoothed, the smoothed value y_i can be calculated as (10). In order to simplify the calculation, the window width is recorded as $2\times l+1$. Due to the limitation of the moving window, the average calculation is only performed on the fixed subset within the window.

$$y_i = \begin{cases} \dfrac{1}{i}\displaystyle\sum_{k=1}^{i} x_k, & i-l<1 \\ \dfrac{1}{n-i+1}\displaystyle\sum_{k=i}^{n} x_k, & i+l>n \\ \dfrac{1}{2l+1}\displaystyle\sum_{k=i-l}^{i+l} x_k, & \text{others} \end{cases} \tag{10}$$

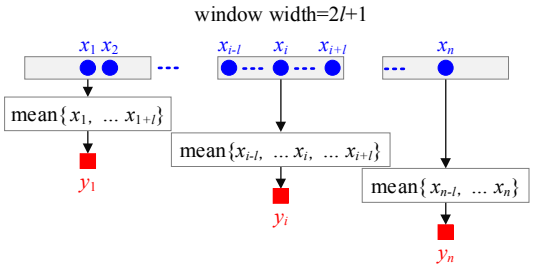

Fig. 2. Diagram of the moving average method.

2) Moving median method (mov-median)

Similar to the mov-average method, the central point and the window width need to be determined first in the mov-median method, as shown in Fig. 3. Then the median value of a set of data in the defined window span is selected to replace the central points. The smoothed can be calculated as

$$y_i = \begin{cases} \text{median}\{x_1 \cdots x_i\}, & i-l>1 \\ \text{median}\{x_i \cdots x_n\}, & i+l>n \\ \text{median}\{x_{i-l} \cdots x_{i+l}\}, & \text{others} \end{cases} \tag{11}$$

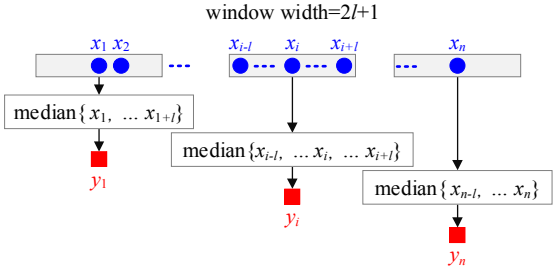

Fig. 3. Diagram of the moving median method.

3) Gaussian filter (gaussian)

Gaussian filter is a linear smoothing method which is suitable to reduce Gaussian noise, and it is a process of weighted average of the raw data. The value of the central data point can be replaced by weighted average of itself and other points in the neighborhood. The detail steps of Gaussian smoothing are as follows:

Step 1: Initializing the parameters of windows width $2\times l+1$ and the variance σ.

Step 2: Calculating the Gaussian weight ω_i for a feature series in the window span $\{x_{i-l},\ldots, x_i, \ldots, x_{i+l}\}$ using the Gaussian function

$$\omega_i = \frac{1}{2\pi\sigma}\exp\left(-\frac{x_i^2}{2\sigma^2}\right) \tag{12}$$

Step 3: Normalizing the weight of the central data x_i as

$$\omega_i = \frac{\omega_i}{\sum_{k=i-l}^{i+l}\omega_k} \tag{13}$$

Step 4: By calculating the weighted sum of row data x_i, the smoothed value y_i is obtained as

$$y_i = \sum_{k=i-l}^{i+l} x_k\omega_k \tag{14}$$

4) Sautzky-Golay filter (s-golay)

S-golay method smooths the data by fitting successive sub-sets of adjacent data points with a low-degree polynomial by linear least squares. Compared with the mov-average method, s-golay can provide better approximation of peaks and valleys. The method is described as follows:

Step 1: Initializing the parameters of windows width $2\times l+1$ and generate the local feature series $\{x_{i-l}, \ldots, x_i, \ldots, x_{i+l}\}$.

Step 2: Using the linear least squares regression to perform the quadratic polynomial fitting on the data in the span.

Step 3: The smoothed value of x_i can be obtained by the regression model, and the above process is repeated across the entire range of data points.

5) Locally weighted scatterplot smoothing (lowess)

The lowess method improves the s-golay using the weight function which gives the most weight to the data points nearest the point of estimation and the least weight to the data points that are furthest away [12]. This method is more robust when there are outliers existing in the dataset, and it can be summarized as follows:

Step 1: Parameters of windows width $2\times l+1$ initialization.

Step 2: The regression weights ω_j for each data in the window span can be obtained by the tricube weight function

$$\omega_j = \begin{cases} \left(1-\left(\dfrac{x_i-x_j}{dd_j}\right)^3\right)^3, & \left|\dfrac{x_i-x_j}{dd_j}\right|<1 \\ 0, & \left|\dfrac{x_i-x_j}{dd_j}\right|\geq 1 \end{cases} \tag{15}$$

where x_i is the central data associated with the response value to be smoothed, x_j are the neighbors of x_i as defined by the window span, dd_j is the maximum distance between x_i and x_j. According to the calculation, the weight is scaled to lie in the range from 0 to 1. Data points outside the span have zero weight and no influence on the fit.

Step 3: Normalizing each weight x_j, then performing the weighted least-squares regression. In lowess smoothing method, the first degree polynomial is used in the regression, thus the optimal parameters can be obtained by minimizing the weighted sum of the squares

$$S = \sum \omega_j\left(a+bx_j-y_j\right)^2 \tag{16}$$

Step 4: The smoothed value of x_i can be obtained by the weighted regression.

6) Robust locally weighted scatterplot smoothing (rlowess)

The rlowess method is more sensitive to the outlier existed in the dataset, and it assigns lower weight to outliers in the regression. When the data outside six mean absolute deviations, the weight will be assigned zero [12]. The difference between lowess and rlowess is step 2, where the median value needs to be found and the weight can be calculated by the 'bisquare' function

$$\omega_j = \begin{cases} \left(1-\left(\dfrac{x_i-x_j}{6s}\right)^2\right)^2, & \left|x_i-x_j\right|<6s \\ 0, & \left|x_i-x_j\right|\geq 6s \end{cases} \tag{17}$$

where s is the median of the residuals in the span.

C. Maching learning-based SOH estimation methods

1) Support vector machine (SVM)

SVM uses kernel technique to map features vectors to high-dimensional space, which is an effective method to deal with nonlinear regression problems [13]. A SVM model is established to capture the nonlinear relationship between smoothed feature and SOH. The objective of SVM is to find the optimal coefficients $\mathbf{w}$ and b on the basis of the following constrained optimization problem,

$$\begin{aligned} \min_{w,b}\ & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ s.t.\ & \mathbf{y}_i-\mathbf{w}^T\cdot\mathbf{x}-b\leq\varepsilon \\ & \mathbf{w}^T\cdot\mathbf{x}+b-\mathbf{y}_i\leq\varepsilon \end{aligned} \tag{18}$$

where $\mathbf{x}\in R^d$ is d-dimensional input features vectors, and $y_i\in R$ is SOH. After solving (18), the SOH estimation function is

$$f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^*-\alpha_i)K(\mathbf{x}_i,\mathbf{x})+b \tag{19}$$

where α_i^* and α_i are Lagrange multipliers, $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function. The radial basis function kernel with the form

of $K(\mathbf{x}_i,\mathbf{x})=\exp\left(-\|\mathbf{x}_i-\mathbf{x}\|^2/2\gamma\right)$ is used because it has advantages in solving nonlinear relationships.

2) Gaussian process regression (GPR)

The Gaussian process is defined as a collection of random variables with joint multivariable Gaussian distribution, and could provide not only the mean value but also the variance of the conditionally expected value of the output [6]. A Gaussian process is expressed as

$$f(\mathbf{x}) \sim N(m(x),k(x,x')) \tag{20}$$

where $m(x)$ and $k(x, x')$ are the mean and covariance functions, respectively. For a regression problem, the output can be modeled as

$$y=f(x)+\varepsilon \tag{21}$$

So the prior distribution can be expressed as

$$\mathbf{y} \sim N(0,K_f(x,x')+\sigma^2\mathbf{I}) \tag{22}$$

By the maximum likelihood method, the hyper parameters σ and f can be obtained. Then the model can be used for prediction and for a new data, the joint prior distribution of $\boldsymbol{y}$ and predicted value $\boldsymbol{y}^*$ can be deduced as

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{y}^*\end{bmatrix} \sim N\left(0,\begin{bmatrix}K_f(x,x)+\sigma^2\mathbf{I} & K_f(x,x^*)\\ K_f(x,x^*)^T & K_f(x^*,x^*)\end{bmatrix}\right) \tag{23}$$

Then the posterior distribution for the a given input is

$$p(\mathbf{y}^*\mid\mathbf{x},\mathbf{y},\mathbf{x}^*)=N\left(\bar{\mathbf{y}}^*,\mathrm{cov}(\mathbf{y}^*)\right) \tag{24}$$

III. Experimental Test

The parameters of the tested $LiFePO_4$ batteries are listed in Table I. As shown in Fig. 4, two battery cells were stored at 50% state of charge (SOC) and 40°C where they underwent calendar aging for a period of 45 months. After each one-month of calendar aging, the current battery capacity was measured at 25°C following a 1C-rate constant current discharging procedure. The SOH is calculated as the ratio between the current available capacity and the initial available capacity, as shown in Fig. 5(a). The battery is considered to reach its end-of-life when its capacity fades by 20% of the initial value. After the battery capacity was measured, these cells were charged at 25°C with a 1C-rate constant current to 20% SOC, 50% SOC and 80% SOC, respectively. At each SOC level, a 33-second pulse test with 4C-rate current was conducted and the voltage responses were used for FE feature extraction. The whole test stopped when the battery reaches its end-of-life criterion, that is, its capacity fades by 20% of the initial value. The collected voltage data are shown in Fig. 5(b) and the corresponding FE features without smoothing are shown in Fig. 5(c). One battery cell (i.e., No. 1) was used for model training and another one (i.e., No. 2) was used for validation. The simulation time (t_{sim}), the root-mean-squared error (*RMSE*), the mean absolute error (*MAE*) and the coefficient of determination *R*-squared (R^2) are used as the indicators to evaluate the performance of each smoothing method, which are defined as:

$$RMSE=\sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T}\left(\hat{SOH}_i-SOH_i\right)^2} \tag{25}$$

$$MAE=\frac{1}{N_T}\sum_{i=1}^{N_T}\left(\frac{\left|\hat{SOH}_i-SOH_i\right|}{SOH_i}\right) \tag{26}$$

$$R^2=1-\frac{SS_{res}}{SS_{tot}}=\frac{\sum_{i=1}^{N_T}(SOH_i-\hat{SOH}_i)^2}{\sum_{i=1}^{N_T}(SOH_i-\bar{SOH}_i)^2} \tag{27}$$

where N_T is the total number of validation data, $\hat{SOH}_i$ and SOH_i is the estimated SOH and the real SOH of the *i*th validation data point, respectively. SS_{res} is the sum of squares of residuals, describes the deviation between the measured points SOH_i and estimated curve $\hat{SOH}_i$, SS_{tot} is the total sum of squares, describes the deviation between the measured points $\hat{SOH}_i$ and their average value $\bar{SOH}_i$.

TABLE I. The Datasheet of the $LiFePO_4$/C Battery.

Item	Value
Nominal voltage	**3.3 V**
Nominal capacity	**2.5 Ah**
Charge voltage	**3.6 V**
Cut-off voltage	**2.0 V**
Maximum continuous charge current	**10 A**
Maximum continuous discharge current	**50 A**

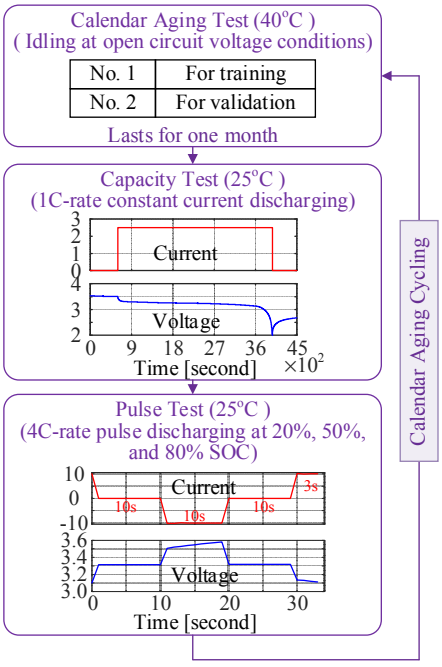

Fig. 4. Flowchart of the test schedules.

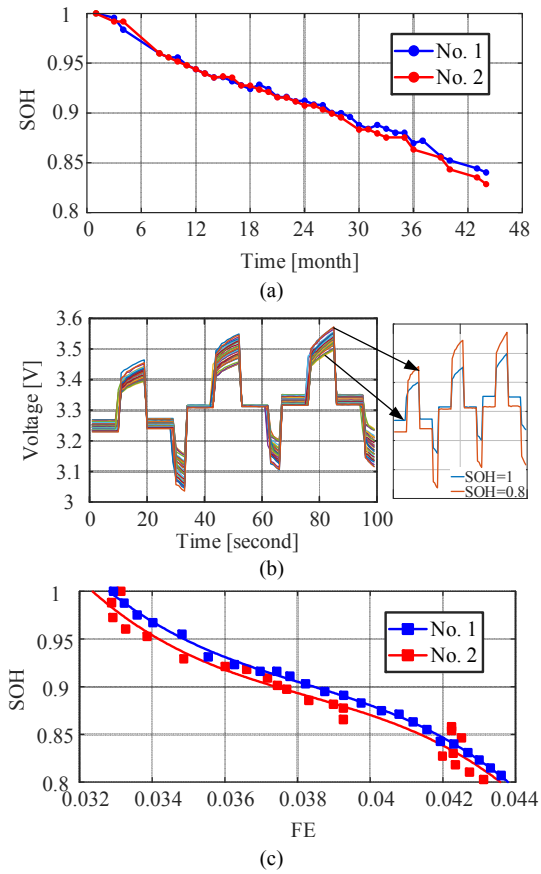

Fig. 5. Results of calendar aging test. (a) SOH curves, (b) Voltage responses during pulse test, (c) FE variation.

IV. SOH ESTIMATION RESULTS

To demonstrate the accuracy and versatility of the established model, the mutual validation method is used in this paper. The data from No. 1 battery is used for SVM/GPR training, while the data from No. 2 battery is for model verification. The SOH estimation results based on SVM algorithm can be seen in Fig. 6 and correspondingly, the comparison of the four performance indicators is shown in Fig. 7. For the GPR algorithm, the results are shown in Fig. 8 and Fig. 9. It can be seen that the estimated SOH fluctuates by a relative large margin around the true value when using the raw feature data to estimate the SOH. The outlier shows up at the fifteenth month and the estimated value of SOH gradually diverges at the end period of aging. However, when the smoothed feature is used, both the MAE and RMSE estimation error decrease as well as the t_{sim}.

According the estimation results based on SVM and GPR, it can be seen that data smoothing has a higher effect on improving the performance of GPR. The errors of GPR-based method are larger than that of SVM-based method when the raw feature is used. The MAE and RMSE for SVM model is 0.009 and 0.016, while these two errors for GPR model is 0.012 and 0.018, respectively. The t_{sim} for GPR model (i.e., 25s) is also longer than that for SVM model (i.e., 10s). However, after pre-processing the data by no matter which smoothing method, GPR model shows a better estimation performance in terms of the accuracy and simulation speed. The MAE and RMSE for SVM are reduced to 0.006, while these two errors for GPR is reduced to 0.003 and 0.005, respectively. The t_{sim} of both methods is reduced to 1s. Among the considered six smoothing methods, the best one is the lowess method. By using lowess for data smoothing, the MAE and RMSE for the SVM model decreases from 0.01 to 0.004 and from 0.02 to 0.005, respectively. The t_{sim} of SVM-based estimation is shortened from 9.56s to 0.98s. While for the GPR model, the MAE and RMSE decreases from 0.01 to 0.002 and from 0.02 to 0.003, respectively. The t_{sim} of GPR-based estimation is shortened from 24.50s to 1.08s.

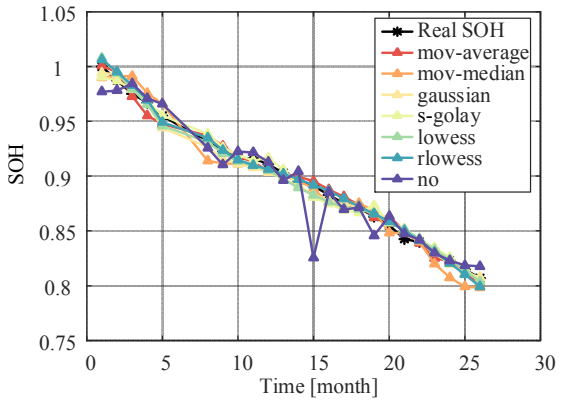

Fig. 6. Estimation results based on SVM.

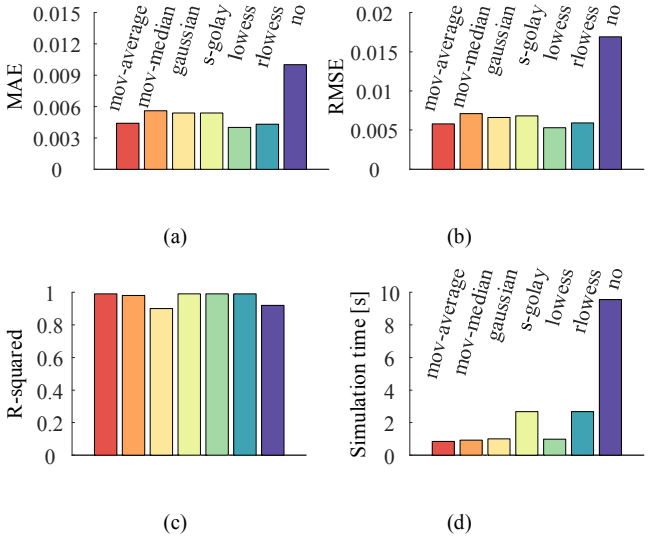

Fig. 7. Comparison results using SVM model. (a) MAE, (b) RMSE, (c) R-squared, (d) Simulation time t_{sim}.

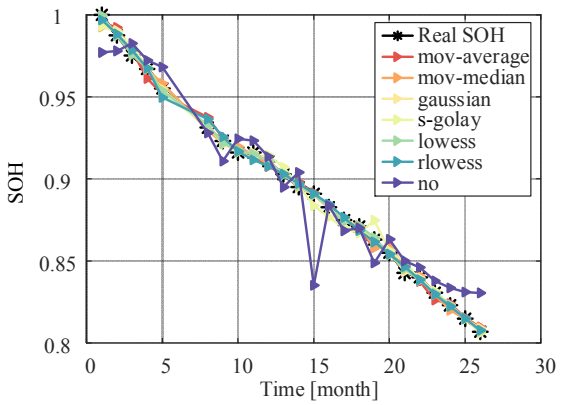

Fig. 8. Estimation results based on GPR.

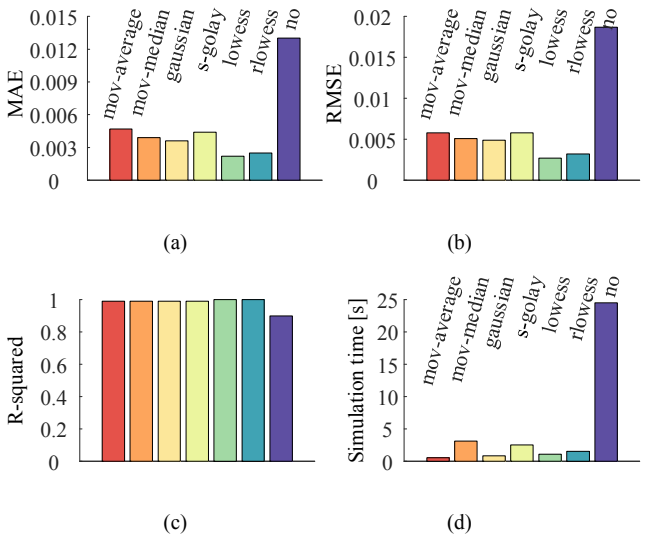

Fig. 9. Comparison results using GPR model. (a) MAE, (b) RMSE, (c) R-squared, (d) Simulation time t_{sim}.

V. Conclusion

This paper studies the effect of data smoothing on the data-driven battery SOH estimation. Six commonly used data smoothing methods are introduced to pre-process the FE feature, which is later used to train the SOH estimation model. As for comparison, SVM and GPR are used to establish the relationship between FE and SOH. Two $LiFePO_4$ batteries are tested under the calendar aging and the FE of the voltage response in the pulse test is used as the feature. The results show that all the six smoothing methods can improve the estimation accuracy and simulation speed to varying degrees. For SVM model, data smoothing helps reduce the estimation error by about 33% (i.e., MAE is reduced from 0.009 to 0.006) and shorten the simulation time by 90% approximately (i.e., t_{sim} is shortened from 10s to 1s). The improvement on the performance of GPR model is more obvious than that of SVM. The estimation error and simulation time are reduced by 75% (i.e., MAE is reduced from 0.012 to 0.003) and 96% (i.e., t_{sim} shortened from 25s to 1s), respectively. In addition, among these methods, the lowess smoothing is the best for improving the performance of FE-based SOH estimation.

References

[1] D. I Stroe, M. Swierczynski, A. Stroe, R. Laerke, P. C. Kjaer and R. Teodorescu, "Degradation behavior of lithium-ion batteries based on lifetime models and field measured frequency regulation mission profile," *IEEE Trans. Ind. Appl.* vol. 52, 2016, pp. 5009-5018.

[2] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renew. Sust. Energ. Rev*. vol. 56, 2016, pp. 572-587.

[3] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *J. Power Sources*, vol. 405, 2018, pp. 18–29.

[4] H. Li, D. Pan and C. L. P. Chen, "Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine," *IEEE Trans. Syst. Man, Cybern. Syst*. vol. 44, 2014, pp. 851-862.

[5] C. Yu, Y. Li, and M. Zhang, "Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting," *Energy Convers. Manage*. vol. 147, 2017, pp. 75-85.

[6] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *J. Power Sources,* vol. 357, 2017, pp. 209-219.

[7] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C. W. Chan, P. van den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Appl. energy,* vol. 232, 2018, pp.197-210.

[8] X Li, C Yuan, Z Wang, "Multi-time-scale framework for prognostic health condition of lithium battery using modified Gaussian process regression and nonlinear regression," *J. Power Sources*, vol. 467, 2020, pp. 228358.

[9] T Liang, L Song, K Shi, "On-board incremental capacity/differential voltage curves acquisition for state of health monitoring of lithium-ion batteries," *in IEEE Int. Conf. Appl. Syst. Invov. (ICASI)*, 2018, pp. 976-979.

[10] W. Chen, Z. Wang, H. Xie and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Trans. Neural Syst. Rehabil. Eng*. vol. 15, 2007, pp. 266-272.

[11] X. Sui, D. I. Stroe, S. He, X. Huang, J. Meng, and R. Teodorescu, "The effect of voltage dataset selection on the accuracy of entropy-based capacity estimation methods for lithium-ion batteries," *Appl. Sci.* vol. 9, 2019, pp. 4170.

[12] R. K. Kimmel, D. E. Booth, and S. E. Booth, "The analysis of outlying data points by robust Locally Weighted Scatter Plot Smooth: a model for the identification of problem banks," *Int. J. Oper. Res*. vol. 7, 2010, pp. 1-15.

[13] J. Tian, R. Xiong and W. Shen, "State of health estimation based on differential temperature for lithium ion batteries," *IEEE Trans. Power Electron*. 2020, Early Access.